journal

why this effort?

Well, I contemplate about a lot of things and it seems that after a point I am not able to keep track of them. I forget some important conclusions, perspectives, schools of thought, jewels of wisdom and such things in various aspects that are fruits of a lot of effort, suffering and pain. And I often have to re-start my train of thought on them, which is tedious and dis-motivating after a point of time.

Therefore this journal is to

1. Write down stuff
   (a) In a simplified, clean, elegant, minimalistic form.
   (b) So that I don't lose it permanently.
   (c) So that I don't have to remember what I don't need to.
   (d) So that I use my brain for thinking rather than storing.
2. Use the power of machine search.
3. Reuse previous knowledge. Don't start over from scratch unncessarily.

The effort will be to

1. Build constructs that are simple, minimal, atomic, composable and excel at one thing.
2. Build meaningful connections and flow among constructs.
3. Keep everything as simple and minimal as possible.
4. Promote recognition rather than re-call.

These ideas are inspired by the UNIX philosophy, my experience in programming and doing things in general.

the process of human learning

The word ‘human’ is important, as mathematics, science, philosophy etc... are after all, human made things and there is nothing ‘absolute’ or ‘exact’ about them. So, there is no reason to be serious or stuck up with anything (especially science). Beliefs lies at the heart of all human made things (even science and the mighty math). As beliefs are not unconditionally correct, nothing should be blindly accepted forever, rather everything better be constantly challenged. This does not mean that nothing should be trusted, that might be even worse.

It is in the **balance** between having trust and challenging it simultaneously lies any **conceivable learning**. If the balance tips left it is called ‘arrogance’, else if it tips right it is called ‘insanity’. If the balance is right it produces a sense of happiness, fun and joy.

I believe that this feeling is what I am after. Science, math, whatever are just classifications of work but the balance is the real deal. **In fact, discovering joy through this balance is the spirit of life itself**.

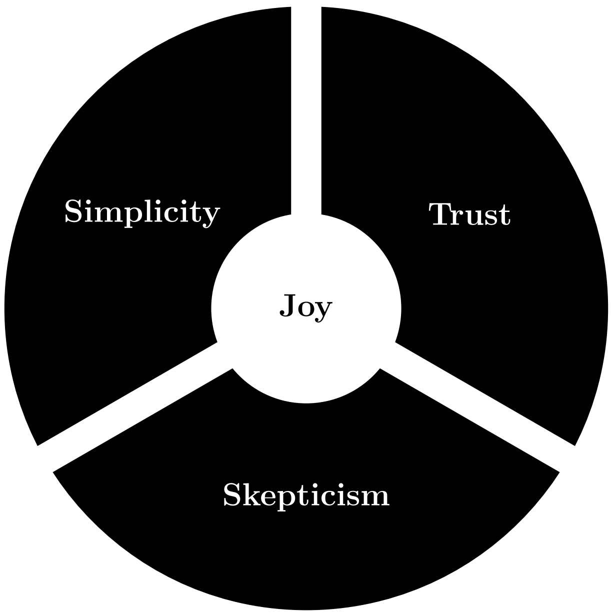


Figure 1: Zen of learning

structure

what?

This is not a re-write of any book. This is the written form of my perspective and understanding of mathematics.

In big picture this text describes a directed acyclic graph (DAG).

Any node of the DAG has a **statement**. A statment is a sentence that is either true or false.

Every statement is associated with a **proof**. All edges from a statement to its parents are associated with the same proof. Given all its parents are true, A statement is true $\longleftrightarrow$ A proof is true. It is illustrated in the figure 2.

As the DAG is constructed manually, the statements are by default in a topological sort. Viz, a statement can not have any statement defined after it as parent, therefore what it states is independent of all such statements.

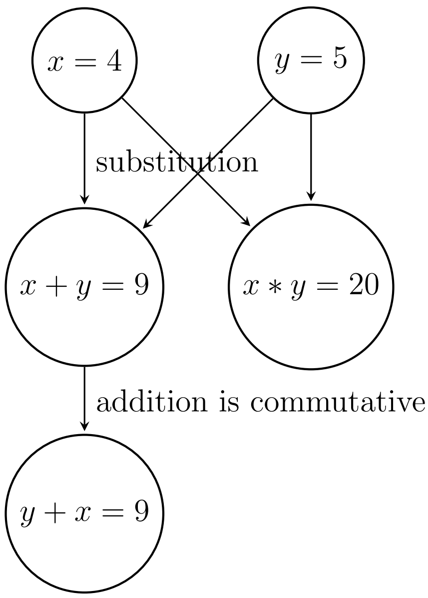


Figure 2: Illustration of structure of this text

why?

Graph because, the hope is that by seeing the bigger picture with connections neatly lay-ed out we might be able to find new things.

Directed because, it is the way mathematics is developed traditionally. Also this way, given a statement, we can get all statements that should be true for the subject statement to be true.

Acyclic because, if there are circles in the statement graph then there is a circular dependency among statements to be true. This makes not much sense to me and therefore acyclic.

also

Statements always have a significance viz. there is a reasonable answer to the question 'why the hell are we even spending time on this?'. I write this for every statement. I avoid purposeless statements.

All definitions in this text are just my convention, which will mostly be same as the general consensus.

Contents

what is mathematics to me?

Statement 1 (Mathematics).

Description: Mathematics is a way of using imagination to solve problems. The essence is that we have a problem and we try to use our imagination to come up with something that solves the problem. There is no further restriction on the definition of math.

Significance: There is also no real boundary b/w normal everyday thinking and mathematics. Just that as thinking becomes more rigourous, it becomes more mathematical. Concepts developed while solving problems like algebra, calculus, statistics etc... are just tools, not the math itself. We can always create new tools and discard old ones.

Proof. Axiom. □

0 parents

basics

Statement 2 (**Identity**).

Description: Two things A and B are identical $\longleftrightarrow$ $\forall context$ there is no distinction b/w them. Denoted by $A \equiv B$.

Two things A and B are non-identical $\longleftrightarrow$ $\forall context$ there is distinction b/w them. Denoted by $A \not\equiv B$.

Significance: Note that this operator is binary, i.e. it takes two operands and produces a boolean. The definition itself is quite simple but I wanted to focus on the word context. For example, a Volkswagen car and a BMW car can be considered identical when the analysis is about classifying objects as cars and kittens. But they can be considered non-identical if we are analyzing the properties of different car brands.

Proof. Axiom. □

0 parents

9 children Order, Enumerated Notation, Set, Belongs To, Proper Subset, Universal Set, Properties of sets, Deterministic Experiment, Non-deterministic Experiment,

Statement 3 (**Equality**).

Description: Two things A and B are equal at a context C $\longleftrightarrow$ there is no distinction b/w them at C. Denoted by $A = B$.

Two things A and B are unequal at a context C $\longleftrightarrow$ there is distinction b/w them at C. Denoted by $A \neq B$.

Significance: Note that this operator is binary, i.e. it takes two operands and produces a boolean. The definition itself is quite simple but I wanted to focus on the word context. For example, if we are talking about the apparant positions of objects in sky to an observer, the sun and moon are equal when it is a complete solar eclipse for her. But they can be considered unequal at all other times.

Proof. Axiom. □

0 parents

0 children

Statement 4 (Order).

Description: Two non-identical [identity] things are ordered $\longleftrightarrow$ there exists an notion that one thing comes before another. If A comes before B then we write $A < B$ or $B > A$.

Significance: There need not be order in all non-identical pairs of things. For example there is no inherent order b/w a rabbit and a horse when listing out all animals. But there is an order when we consider the heights of two students when listing out heights of students in a class.

Proof. Proof by definition □

1 parents Identity,

1 children Ordered Set,

Statement 5 (Zero).

Description: The notion of absence. Symbolically written as 0. Can be considered as a number.

Significance: For example in case of counting it can represent the case when there is no balls left in an urn.

Proof. Axiom. □

0 parents

3 children Null Set, Whole Numbers, Standard Vector,

Statement 6 (Number).

Description: A number is a notion to count or to label or measure things.

Significance: Some uses of numbers are

1. Count number of apples in a bag - There are 37 apples in a bag
2. Measure the length of a pencil - The length of the pencil is 3.1468 cm
3. Label each participant in a marathon - The participant number 10714 won the marathon

Proof. Axiom. □

0 parents

2 children Enumerated Notation, Natural Numbers,

Statement 7 (**Notation**).

Description: A shortform representation of a tool or idea. The representation is short doesn't mean that information is omitted, rather it motivates to present information in a more concise and simplified manner. A good notation has complete information, is transparent in meaning, readable and un-ambiguous.

Significance: For instance, when writing a vector of a position g, writing P_g is a bad notation as it has no information about the coordiante system used. P_g^Λ where Λ is the coordinate system is a good notation.

Proof. Axiom. □

0 parents

0 children

Statement 8 (**Enumerated Notation**).

Description: To name each element in a collection such that each elements name is non-identical [identity] to all others, we arrange each element in some way and use the position [number] as the subscript of that element. This is called enumerated notation. For a collection having n elements the enumerated notation would be $\{e_1, e_2, e_3, ...e_n\}$.

- If n = 0, it expands to $\{\}$.
- If n = 1, it expands to $\{e_1\}$
- If n = 2, it expands to $\{e_1, e_2\}$
- ...
- If n = n, it expands to $\{e_1, e_2, ...e_n\}$

Significance: This way we can refer to each element uniquely. This comes in handy all the time.

Proof. Proof by definition □

2 parents Identity, Number,

1 children Set,

sets

Statement 9 (Set)**.**

Description: A set S is a collection of elements from which any two elements picked without replacement are non-identical [identity]. If the elements are denoted by $e_1, e_2, e_3, ...e_n$ the set is denoted by $S \equiv \{e_1, e_2, e_3, ...e_n\}$ [enumerated notation].

Significance: To represent collection of differently colored balls in an urn, etc...

Proof. Proof by definition □

2 parents Identity, Enumerated Notation,

23 children Belongs To, Subset, Superset, Proper Subset, Proper Superset, Union, Intersection, Complement, Universal Set, Null Set, Ordered Set, Properties of sets, Partition, Function, Whole Numbers, Linearly Independent Set, Span, Basis, Standard basis, Space, State Vector Space, Outcome Space, Event Space,

Statement 10 (Belongs To)**.**

Description: An element λ belongs to a [set] S $\longleftrightarrow \exists! \tau \in S \mid \lambda \equiv \tau$ [identity]. Denoted by $\lambda \in S$.

An element λ does not belong to a S $\longleftrightarrow \nexists \tau \in S \mid \lambda \equiv \tau$. Denoted by $\lambda \notin S$.

Significance: To indicate whether an element belongs to a set or not. There cannot be more than one element in S that is identical to λ anyways due to the definition of a set.

Proof. Proof by definition □

2 parents Set, Identity,

3 children Subset, Complement, Function,

Statement 11 (Subset)**.**

Description: A [set] E is a subset of set F $\longleftrightarrow \forall e \in E \implies e \in F$. Denoted by $E \subseteq F$ [belongs to].

Significance: Can be used to represent useful pieces of a set.

Proof. Proof by definition □

2 parents Set, Belongs To,

6 children Superset, Proper Superset, Partition, Coordinate Axis, Straight Space, Event,

Statement 12 (Superset).

Description: A [set] E is a superset of set F $\longleftrightarrow F \subseteq E$ [subset]. Denoted by $E \supseteq F$.
Significance: The inverse of subset.

Proof. Proof by definition □

2 parents Set, Subset,

0 children

Statement 13 (Proper Subset).

Description: A [set] E is a proper subset of set F $\longleftrightarrow E \subseteq F$ and $E \not\equiv F$ [identity]. Denoted by $E \subset F$.
Significance: Can be used to represent useful pieces of a set.

Proof. Proof by definition □

2 parents Set, Identity,

0 children

Statement 14 (Proper Superset).

Description: A [set] E is a superset of set F $\longleftrightarrow F \subset E$ [subset]. Denoted by $E \supset F$.
Significance: The inverse of subset.

Proof. Proof by definition □

2 parents Set, Subset,

0 children

Statement 15 (Union).

Description: A union of two [set]s A and B is a set which contains every element in A, every element in B, with repetitions removed, if any. Denoted by $A \cup B$.
Significance: A basic operator to combine sets.

Proof. Proof by definition □

1 parents Set,

2 children Universal Set, Whole Numbers,

Statement 16 (Intersection).

Description: An intersection of two [set]s A and B is a set which contains every element that is common to A and B. Denoted by $A \cap B$.

Significance: A basic operator to combine sets.

Proof. Proof by definition □

1 parents Set,

0 children

Statement 17 (Complement).

Description: A complement of a [set] A is a set which contains all elements which do not [belongs to] the set A itself. Denoted by $\overline{A}$

Significance: A basic unary operator on a set.

Proof. Proof by definition □

2 parents Set, Belongs To,

1 children Universal Set,

Statement 18 (Universal Set).

Description: A universal set of a [set] A is a set which is identical to $A \cup \overline{A}$ [identity] [union] [complement].

Significance: A representation of a whole. This is highly context dependent i.e. it depends on what one believes to be in the complement. A proper superset of universal set doesn't exist. One can argue to just put some more elements to build a superset, but it makes things complicated. If there is ever a real need for such a construct (dynamic universal set!!!???), we can build it then.

Proof. Proof by definition □

4 parents Set, Identity, Union, Complement,

2 children Properties of sets, Outcome Space,

Statement 19 (Null Set).

Description: A null set is a [set] which has [zero] elements in it. Denoted by ϕ.

Significance: A representation for nothingness. The reason for my take is because even null set is just a mathematical tool. There is a great debate whether there is a single null set or many null sets. My take on it is that it depends on the context.

Proof. Proof by definition □

2 parents Set, Zero,

7 children Properties of sets, Partition, Mutually Exclusive Events, Properties of probability, Conditional Probability, Total Probability Theorem, Bayes Theorem,

Statement 20 (Ordered Set).

Description: An ordered set is a [set] where there exists an [order] b/w every pair of elements in it.
Significance: Useful to represent set of heights of students in a class, etc...

Proof. Proof by definition □

2 parents Set, Order,

3 children Continuous Set, Contiguous Set, Natural Numbers,

Statement 21 (Continuous Set).

Description: An [ordered set] which has no discrete separation b/w any two of its elements is a continuous set.
Significance: To represent sets like the set of all wavelengths in the visible spectrum of light

Proof. Proof by definition □

1 parents Ordered Set,

0 children

Statement 22 (Contiguous Set).

Description: An [ordered set] which has a discrete separation b/w any two of its elements is a contiguous set.
Significance: To represent colors in the rainbow in order [Violet, Indigo, Blue, Green, Yellow, Orange, Red]

Proof. Proof by definition □

1 parents Ordered Set,

0 children

Statement 23 (Properties of sets).

Description: The following list enumerates some properties of [set]s. Here S denotes [universal set] and ϕ denotes [null set] [identity].

1. $A \cup B \equiv B \cup A$ (Commutative)
2. $A \cap B \equiv B \cap A$ (Commutative)
3. $(A \cup B) \cup C \equiv A \cup (B \cup C)$ (Associative)
4. $(A \cap B) \cap C \equiv A \cap (B \cap C)$ (Associative)
5. $A \cup A \equiv A$ (Idompotent)
6. $A \cap A \equiv A$ (Idompotent)

7. $A \cup S \equiv S \cup A \equiv S$
8. $A \cap S \equiv S \cap A \equiv A$
9. $A \cup \phi \equiv \phi \cup A \equiv A$
10. $A \cap \phi \equiv \phi \cap A \equiv \phi$
11. $A \cup \overline{A} \equiv S$ (Definition of universal set)
12. $A \cap \overline{A} \equiv \phi$
13. $A \subseteq B \longleftrightarrow A \cup B \equiv B$
14. $A \subseteq B \longleftrightarrow A \cap B \equiv A$
15. $A \cap (B \cup C) \equiv (A \cap B) \cup (A \cap C)$ (Distributive)
16. $A \cup (B \cap C) \equiv (A \cup B) \cap (A \cup C)$ (Distributive)

Significance: Some basic relations b/w sets so we can play with them.

Proof. Proof by using definitions of parent statements. □

4 parents Set, Universal Set, Null Set, Identity,

0 children

Statement 24 (Partition).

Description: A partition of a [set] S is a set Λ of [subset]s of S such that

1. Intersection of any two elements of $\Lambda \equiv \phi$ [null set]
2. Union of all elements of $\Lambda \equiv$ S

Significance: A neat way to divide a set.

Proof. Proof by definition □

3 parents Set, Subset, Null Set,

2 children Total Probability Theorem, Bayes Theorem,

functions

Statement 25 (Function).

Description: A function Λ on a [set] called its domain D is a mapping from each element α that [belongs to] D to an element β in a new set C called its co-domain. This mapping exists for all elements in the domain and each element in domain is mapped to only one element in co-domain. For each element $\alpha \in D$ the mapping β is called image of α. The set of images for all elements the domain is called range R of the function Λ. Function is denoted by $\Lambda : D \to C$. The image of α is denoted by $\Lambda(\alpha) \forall \alpha \in D$.

Significance: A basic tool for moving b/w sets. Note that two elements on domain can be mapped to same element in co-domain. Also range $\subseteq$ co-domain.

Proof. Proof by definition □

2 parents Set, Belongs To,

4 children Derivative, Integral, Fundamental theorem of Calculus, Random Variable,

Statement 26 (Injection).

Description:**TODO**
Significance:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 27 (Surjection).

Description:**TODO**
Significance:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 28 (**Bijection**).

Description:**TODO**
Significance:**TODO**

Proof. Axiom. □

0 parents

2 children Coordinate System, Linear Transformation,

numbers

Statement 29 (Natural Numbers).

Description: The [ordered set] of all numbers [number] used for counting and ordering is called the set of natural numbers. It is denoted by $\mathbb{N}$.

Significance: Encompasses all possible counts.

Proof. Proof by definition □

2 parents Ordered Set, Number,

2 children Whole Numbers, Deterministic Experiment,

Statement 30 (Whole Numbers).

Description: The [union] of [natural numbers] and [set] which contains only [zero] . It is denoted by $\mathbb{W}$.

Significance: The name whole is given because we are gonna define numbers that will represent non-whole entities.

Proof. Proof by definition □

4 parents Union, Natural Numbers, Set, Zero,

0 children

Statement 31 (Additive Inverse).

Description:**TODO**
Significance:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 32 (Integers).

Description:**TODO**
Significance:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 33 (Multiplicative Inverse).

> **Description**:**TODO**
> **Significance**:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 34 (Rational Numbers).

> **Description**:**TODO**
> **Significance**:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 35 (Irrational Numbers).

> **Description**:**TODO**
> **Significance**:**TODO**

Proof. Axiom. □

0 parents

0 children

Statement 36 (Real Numbers).

> **Description**:**TODO**
> **Significance**:**TODO**

Proof. Axiom. □

0 parents

5 children Vector, Vector Scaling, Linear Combination, Dot Product, Straight Space,

vectors

Statement 37 (**Vector**).

Description: A vector can be defined as a collection of [real numbers] (also called elements) where the arrangement in which elements appear matters. For n elements a vector can be written as

$$V \equiv e$$

Significance: The motivation behind vectors is to view a group of entities as a single entity. By viewing group of multiple entities as a single entity, a more abstract concept can be created, where instead of applying the same operation to each and every element, again and again, we apply the same operation to the whole group entity at once i.e. vector. A simple use case would be say when you need to update marks of all students to a 100 scale from a 10 scale.

Proof. Proof by definition □

1 parents Real Numbers,

15 children Vector Dimension, Vector Addition, Vector Scaling, Linear Combination, Linearly Independent Set, Span, Basis, Standard Vector, Standard basis, L2 Norm, Dot Product, Angle between Vectors, Perpendicular Vectors, State Vector, Linear Transformation,

Statement 38 (**Vector Dimension**).

Description: The dimension of a [vector] is just the count of elements in it. A vector with dimension N can be called N dimensional vector, written as ND vector.

Significance: A name for size of vector. Note that we are not considering collections with zero elements as vectors as it does not seem to be useful. The concept is illustrated in the table 6.1.

Vector name	Value	Dimension
V_1	(1)	1
V_2	$(\sqrt{2})$	1
V_3	(-100, $\sqrt{3}$)	2
V_4	(0, 0.1)	2
V_5	(0, 0, 0)	3
V_6	(0, 1, 2, 3)	4

Table 6.1: Dimensions of Vectors

Proof. Proof by definition □

1 parents Vector,

3 children Vector Addition, Vector Scaling, Linear Combination,

Statement 39 (**Vector Addition**).

Description: The addition of two ND [vector]s

$$A \equiv a$$

and

$$B \equiv b$$

is a new ND vector with each element as sum of corresponding elements in A and B. [vector dimension]. Denoted by

$$A + B \equiv ab$$

.

Significance: This is to re-enforce the notion of applying operation to the group entity rather than each and every element repeatedly. Otherwise there is no reason not to define the addition rule as something arbitrary like

$$xy + ab \equiv xa + yb^2 \frac{b - y}{x + a}.$$

Proof. Proof by definition □

2 parents Vector, Vector Dimension,

1 children Linear Combination,

Statement 40 (**Vector Scaling**).

Description: The scaling of an ND [vector]

$$A \equiv a$$

[vector dimension] with a real number [real numbers] λ produces a new ND vector with each element as product of corresponding elements in A and λ. Denoted by

$$\lambda * A \equiv \lambda a$$

.

Significance: Same story. Here λ is called (surprise surprise) a scalar, a mysterious word that crept in the subject of vectors suddenly becomes not so mysterious after knowing its name is given to it by the work it does.

Proof. Proof by definition □

3 parents Vector, Vector Dimension, Real Numbers,

2 children Linear Combination, Coordinate Axis,

Statement 41 (**Linear Combination**)**.**

Description: Linearly combination two vectors ND [vector]s A and B [vector dimension] means first scaling each vector with a real number [real numbers] and adding resultant vectors [vector scaling] [vector addition]. The produced vector is also N dimensional. Denoted by $\alpha * A + \beta * B$, where α and β are real numbers.

Significance: This is nothing new, just a word to combine one scaling and one adding operations. A higher level construct to play with. What is so linear about it? Well the word linear is given because there can exist other types of combinations like quadratic combination where before scaling the entities are squared, polynomial combination, exponential combination, logarithmic combination ... The most simplest of them all seems to be linear combination.

Proof. Proof by definition □

5 parents Vector, Vector Dimension, Real Numbers, Vector Scaling, Vector Addition,

4 children Linearly Independent Set, Span, Basis, Linear Transformation,

Statement 42 (**Linearly Independent Set**)**.**

Description: A linearly independent set is a [set] of [vector]s B of cardinality > 1, such that $\forall b_i \in B, b_i \not\equiv$ [linear combination] of all other vectors in B.

A set of vectors in which at least one element $\equiv$ a linear combination of all other elements is called a linearly dependent set of vectors.

Significance: Vectors in a set with this property are sort of independent from others as in they cannot be built by other vectors using any simple operation available for vectors. Yes! addition and scaling are all the available simple operations for vectors. Mind blown! Well you might scream matrices. But matrices were created just for that purpose, to build a vector that is otherwise not constructable using addition and scaling.

Note that there cannot be a set of vectors in which one element $b_d \equiv$ linear combination of all other elements and another element $b_i \not\equiv$ linear combination of all other elements. This is because given the first identity, we can rearrange terms to get second identity.

Proof. Proof by definition □

3 parents Set, Vector, Linear Combination,

1 children Basis,

Statement 43 (**Span**)**.**

Description: The span of a [set] of [vector]s B is a set of vectors S in which each vector $\equiv$ [linear combination] of all vectors in B.

Significance: A tool to give a sense of spread/cover of a set of vectors.

Proof. Proof by definition □

3 parents Set, Vector, Linear Combination,

0 children

Statement 44 (Basis).

Description: A basis of a [set] of [vector]s S is a [linearly independent set] of vectors B, such that each vector in S $\equiv$ [linear combination] of all vectors in B.

Significance: Sort of an inverse of span. Note that for a given set of vectors there can be multiple basis sets. For instance the set of all 2D vectors has the following sets as basis sets.

$$\{10, 01\}, \{20, 02\}, \{20, 12\}$$

By using the word linearly independent we are sort of putting a restriction on number of vectors in the set.

Proof. Proof by definition □

4 parents Set, Vector, Linearly Independent Set, Linear Combination,

0 children

Statement 45 (Standard Vector).

Description: A standard vector is a [vector] in which value of one and only one element is one. If any elements are remaining their value is [zero].

Significance: An standard vector is a simple unit vector. For example

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \begin{bmatrix}1\\0\\0\end{bmatrix} \begin{bmatrix}1\end{bmatrix}$$

are instances of standard vector. It can be used as a building block for creating useful tools.

Proof. Proof by definition □

2 parents Vector, Zero,

2 children Standard basis, Coordinate Axis,

Statement 46 (Standard basis).

Description: The standard basis of a [set] of N-dimensional [vector]s V is the set of all [standard vector]s of dimension N.

Significance: Can be used to construct coordinate axes and thus coordinate system. As each vector in standard basis belongs to one and only one coordinate axis, any point in the state vector space $\equiv$ linear combination of all vectors in standard basis. For example a set of 4D vectors have a standard basis as the following

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix} \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

Proof. Proof by definition □

3 parents Set, Vector, Standard Vector,

2 children Linear Transformation, Rotation Transformation,

Statement 47 (**L2 Norm**).

Description: The L2 norm of a [vector]

$$V \equiv e$$

is defined as the square root of sum of squares of elements of V. Denoted by

$$V_2 \equiv \sqrt{\sum_{i=1}^{n} e_i^2}$$

.

Significance: A simple and natural way to get a sense of size of vector. This norm is also called Euclidean norm.

Proof. Proof by definition □

1 parents Vector,

2 children Angle between Vectors, Rotation Transformation,

Statement 48 (**Dot Product**).

Description: The dot product of N-dimensional two [vector]s

$$A \equiv aB \equiv b$$

is a real number [real numbers] $\eta \equiv$ sum of element-wise products of A and B. Denoted by

$$A.B \equiv \sum_{i=1}^{n}(a_i * b_i)$$

Significance: A tool to give a sense of similarity of vectors.

Proof. Proof by definition □

2 parents Vector, Real Numbers,

1 children Angle between Vectors,

Statement 49 (**Angle between Vectors**).

Description: The angle between two N-dimensional [vector]s A and B is a real number $\theta \equiv$ arc cosine of [dot product] of vectors divided by [L2 norm] of both A and B. Denoted by

$$\angle(A, B) \equiv arccos(\frac{A.B}{A_2 * B_2})$$

Significance: A tool give a sense of orientation difference b/w the vectors.

Proof. Proof by definition □

3 parents Vector, Dot Product, L2 Norm,

1 children Perpendicular Vectors,

Statement 50 (Perpendicular Vectors).

Description: Two N-dimensional [vector]s A and B are said to be perpendicular vectors $\longleftrightarrow$ the cosine of [angle between vectors] $\equiv 0$.

Significance: Gives a sense of independence b/w two vectors.

In vectors of dimensions 2 and 3, we visualize this by drawing what we usually draw for coordinate systems. This visualization is just a choice. We choose that because the independence is visualized in the sense of projections. In 1D there are only one axis and in 4D and more there is no easy, simple and useful way to visualize perpendicular axes yet. So there is no need to visualize them in 4D or more dimensional spaces.

Proof. Proof by definition □

2 parents Vector, Angle between Vectors,

1 children Rotation Transformation,

coordinate system

Statement 51 (Point).

Description: A point represents state of an object. A point does not change, so there is no notion of moving a point. Objects can change states and hence can go from one point to another.

Significance: A basic tool for representing state of objects. For instance a point can represent position, temperature, velocity, pressure etc...

Proof. Axiom. □

0 parents

6 children Space, State Vector, Coordinate System, Coordinate Axis, Straight Space, Position Space,

Statement 52 (Space).

Description: A space is a [set] of [point]s.

Significance: A tool to group all states considered in the context. As point does not change, space does not change.

Proof. Proof by definition □

2 parents Set, Point,

7 children Coordinate System, Coordinate Axis, Straight Space, Position Space, Linear Transformation, Rotation Transformation, Definiteness,

Statement 53 (State Vector).

Description: A state vector is a [vector] that represents a [point].

Significance: This is not a redundant definition (with point). Although a point never changes, the representation of that point can change. That is why we define both point and state vector.

Proof. Proof by definition □

2 parents Vector, Point,

3 children State Vector Space, Coordinate Axis, Straight Space,

Statement 54 (State Vector Space).

Description: A state vector space is a [set] of all possible N-dimensional [state vector]s.

Significance: Used to represent a collection of state vectors. Mainly created for constructing coordinate system.

Proof. Proof by definition □

2 parents Set, State Vector,

3 children Coordinate System, Linear Transformation, Rotation Transformation,

Statement 55 (Coordinate System).

Description: A coordinate system Λ is a [bijection] mapping from [space] G to an N-dimensional [state vector space] V.

Denoted by $\Lambda : G \to V$. $\forall g \in G, \Lambda(g)$ (i.e. the image of [point] g under coordinate system Λ) can be denoted by P_g^{Λ}.

Significance: We choose state vector space to represent a space G because, by definition any v $\in V \equiv$ linear combination of all vectors of standard basis. Therefore we only need standard basis to describe any vector in V, nothing else.

Sets up a representation for every point in space uniquely i.e. such that any point's representation is different from all others. Defining it like this removes all untagible invisible evil chains that are otherwise tied to the concept, thus providing a clean, elegant and concrete description.

Here we make sort of a first proper connection from points and spaces to vectors. Note that when we talk about dimensionality here, it is the property of vectors not the space itself. Thus space and dimensionality are decoupled, this means a space can be represented by different dimensional vectors.

Proof. Proof by definition □

4 parents Bijection, Space, State Vector Space, Point,

4 children Coordinate Axis, Straight Space, Linear Transformation, Rotation Transformation,

Statement 56 (Coordinate Axis).

Description: A coordinate axis of a [coordinate system] Λ acting on a [space] G is a [subset] of G, say A such that the [state vector] of any [point] in A $\equiv$ a scaled version of an [standard vector] B [vector scaling].

Significance: For a coordinate system mapping to an N-dimensional state vector set, there are N coordinate axes. Due to the virtue of state vectors on coordinate axis being scaled standard vectors, any linear combination of two vectors on a coordinate axis lies on the same coordinate axis. It has no connection with other coordinate axes. This is a good thing because it makes things simple, decoupled and non-clumsy.

A coordinate axis is a simple way to disassemble a coordinate system. Any point in the state vector space ≡ linear combination of some point on each coordinate axis.

Misconception busting:

1. There is no need to name the axes X, Y ...
2. There is no need for the axes to be perpendicular.
3. No need for 1 unit along first axis to be same as 1 unit along second axis.
4. No need for units of first axis to be same as the second one

Proof. Proof by definition □

7 parents Coordinate System, Space, Subset, State Vector, Point, Standard Vector, Vector Scaling,

0 children

Statement 57 (Straight Space).

Description: A straight space under a [coordinate system] Λ acting on a [space] G is a [subset] of G, say S such that the [state vector] of any [point] in S $\equiv o + t * d$ where $o, d \in G$ and t ∈ [real numbers].

Significance: If the space is finite, it is called a line segement otherwise a line. We can see the importance of the operation of linear combination here, as it is the defining operation for lines and line segments. Generally denoted by drawing ↔ on paper.

Proof. Proof by definition □

6 parents Coordinate System, Space, Subset, State Vector, Point, Real Numbers,

0 children

Statement 58 (Position Space).

Description: A [space] where each [point] represents a position of an object is called a position space. The coordinate system that acts on it is called position coordinate system.

Significance: Used to denote positions of objects in the real world. Positions in real world are generally mapped to a 3D vector set. This mapping is just based on the belief that the real world is in 3D. But as we believe, beliefs are not final and should be challanged constantly and consciously.

For example positions can be mapped to 4D vector set, the new dimension being the time. Similarly more dimensions can represent more information. Or the real world might contain something more that we have not yet observed properly which cannot be represented using a 3D vector set.

Whatever the real world might be, we can always imagine a space represented by 3D vector set and do operations on it.

Proof. Proof by definition □

2 parents Space, Point,

linear algebra

Statement 59 (Linear Transformation).

Description: A linear transformation on a [space] G is a [bijection] mapping from a [coordinate system] Λ on G mapping to a [state vector space] V to a coordinate system Ψ on same G mapping to state vector space W of same dimension as V, such that each [vector] in [standard basis] of W is a [linear combination] of each vector in standard basis of V.

Significance: A simple tool for moving b/w representations of space. Some popular examples of linear transformations are rotation, scale, sheer.

For instance in a 2D state vector space, if standard basis of W say $\hat{o}, \hat{p}$ are linear combinations of standard basis of V say $\hat{d}, \hat{f}$ as follows

$$\hat{o} = 3 * \hat{d} + 2 * \hat{f}$$

$$\hat{p} = 0 * \hat{d} + 2 * \hat{f}$$

Note that

1. In V $\hat{o}, \hat{p}$ represent $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ repsectively.
2. In W $\hat{d}, \hat{f}$ represent $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ repsectively.

But V and W are different representations of the same space. Therfore,

1. $\hat{o}$ represents a same vector as $\hat{d}$. So they are indentical in this context.
2. $\hat{o}$ represents a different point than $\hat{d}$. So they are non-identical in this context.

Therefore when we say two vectors are equal we should be careful about the context of the discussion.

Suppose a point η is represented by $\begin{bmatrix} 4 \\ 10 \end{bmatrix}$ in Ψ. That means P is represented by $4 * \hat{o} + 10 * \hat{p}$. Now to get representation of the same point η in Λ (remember points do not change with anything) we just substitute the expressions of $\hat{o}, \hat{p}$ in terms of $\hat{d}, \hat{f}$ as given above, i.e. $4 * (3 * \hat{d} + 2 * \hat{f}) + 10 * (0 * \hat{d} + 2 * \hat{f})$ which gives $12 * \hat{d} + 28 * \hat{f}$. This can be written as

$$4 * \begin{bmatrix} 3 \\ 2 \end{bmatrix} + 10 * \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

which can be further packaged as a matrix multiplication.

$$\begin{bmatrix} 3 & 0 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix}$$

This is why the matrix multiplication was invented. Mind Blown!! Therefore columns in a matrix can be seen as the representations of standard basis of new coordinate system in the old one.

Therefore the linear tranform can be seen as a way to do the following things

1. Given a transformed coordinate system i.e. given standard basis of new coordinate system in terms of the old one, we can get the representation for the same point in two different systems.
2. Visualize moving a vector λ representing a point A in Λ to the point B represented in Ψ by ψ which is numerically equal to λ.

Donot mix these two.

Misconceptions busting. The following are true statements.

1. Rotation is not a property of a body, it is a property of a pair of coordinate systems. One of the coordinate systems can be attached to a body.
2. Rotation is not some complicated thing, it is just a linear transformation. The columns of the matrix representing rotation transform are just the places where the new standard basis land in old coordinate system.
3. There is no such thing as multiplying a vector on its right by a matrix (contrary to left side which we can use for linear transform). Dimensions don't match for multiplication to occur.
4. Linear transformation is not about raw numbers, the idea comes from visualing space and trying to play with it.

Proof. Proof by definition □

7 parents Space, Bijection, Coordinate System, State Vector Space, Vector, Standard basis, Linear Combination,

2 children Rotation Transformation, Definiteness,

Statement 60 (Rotation Transformation).

Description: A rotation transformation is a [linear transformation] on a [space] G from [coordinate system] Λ with [state vector space] V to coordinate system Ψ with state vector space W, in which each vector in [standard basis] of W has [L2 norm] of 1 and is perpendicular to all others [perpendicular vectors]

Significance: Defines a rotation operation. The matrix formed by such transformation is called rotation transformation matrix or rotation matrix for short. The standard basis of W forms an orthonormal basis.

In 3D the number of degrees of freedom for a rotation matrix is 3 even though we have 9 elements in a 3x3 matrix. This is because for a 3D unit vector we have two degrees of freedom (the third one is fixed when we choose first two due to length constraint). Therefore by choosing the unit vector value of first axis of new coordinate system in the old coordinate system we used 2 degrees of freedom. Given this vector we have to choose a perpendicular vector i.e. any vector in plane perpendicular to it. Choosing this will require 1 degree of freedom which is the angle of vector in that plane. The final vector is just the cross product of first two which uses 0 degrees of freedom.

Say there is a state vector v. After a rotation tranformation the same state vector v in new coordinate system is represented by state vector u in old one. Two vectors form a plane. So u, v form a plane P that passes through the origin. Every plane has one and only one normal, P also has one and only one normal n. The rotation transformation matrix has a determinant $\equiv 1$ because it is a orthonormal matrix. Therefore the length (L2 norms) of both vectors u, v are identical. In the plane P we have angle b/w u, v say θ. Therefore the whole rotation transformation can be seen as a rotating the state vector v about one and only one axis n by angle θ to reach u. There you have the good old way to thinking about rotation. Mind blown!!

Misconceptions busting. The following are true statements.

1. Rotation is not a property of a body, it is a property of a pair of coordinate systems. One of the coordinate systems can be attached to a body.
2. Rotation is no some complicated thing, it is just a linear transformation. The columns of the matrix representing rotation transform are just the places where the new standard basis land in old coordinate system.
3. In 3D, every rotation always has one and only one axis of rotation which passes through origin. Rotation about an axis not passing through origin is not defined.
4. Similarly, in 2D, every rotation is always about origin. Rotation about a point which is not an origin is not defined.
5. Although rotation is a linear transform, the result of rotation may not be linear in terms of the angle of rotation.

$$\begin{bmatrix} a \\ b \end{bmatrix} \equiv \begin{bmatrix} c\theta & -s\theta \\ s\theta & c\theta \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} \equiv \begin{bmatrix} xc\theta - ys\theta \\ xs\theta + yc\theta \end{bmatrix}$$

For example here the result $[a, b]^T$ is linear in terms of $c\theta, s\theta$ and not θ even though entire rotation matrix can be generated only using θ. A bit misleading indeed.

Proof. Proof by definition □

7 parents Linear Transformation, Space, Coordinate System, State Vector Space, Standard basis, L2 Norm, Perpendicular Vectors,

0 children

Statement 61 (Definiteness).

Description: Let a square matrix M represent a [linear transformation] on a [space] G. If

1. $z^T Mz > 0$ for any non-zero vector $z \longleftrightarrow$ it (transformation) is called positive definite.
2. $z^T Mz < 0$ for any non-zero vector $z \longleftrightarrow$ it is called negative definite.
3. $z^T Mz \geq 0$ for any non-zero vector $z \longleftrightarrow$ it is called postive semi-definite.
4. $z^T Mz \leq 0$ for any non-zero vector $z \longleftrightarrow$ it is called negative semi-definite.
5. else it is called indefinite.

Significance: A neat way to classify transformations. In the expression $z^T Mz$, Mz is the transformed vector of z. Therefore the entire expression is the dot product of the vector to its transformed vector. Hence a positive definite transformation $\longleftrightarrow$ dot product of any vector with its transformed one > 0. Similarly for others.

Proof. Proof by definition □

2 parents Linear Transformation, Space,

0 children **TODO**

1. composition of linear transforms, composition as product of matrices, building matrix as tool
2. proof of composition of linear transform is a linear transform, algebraically and geometrically
3. misconceptions: ~~animating vectors~~ vs ~~animating coordinate frames~~ don't try to animate (lerp) anything
4. latest frame vs initial frame
5. representation of rotation: rot mat, quaternions, axis-angle, eular, roll pitch yaw, gimbal lock
6. determinant, descriminant whatever
7. cross product
8. inverses
9. column spaces, rank
10. null spaces
11. eigen stuff, eigen basis
12. orthogonal, orthonormal
13. diagnoal
14. quaternions
15. SVD, other easy useful decompositions
16. quadratic systems using matrices like X^TAX.
17. derivatives of matrices, matrix products like X^TX.

TODO

1. multiplication of quaternions is not commutative
2. rot mat $< -$‘ $>$ eular angle $< - >$ rpy
3. what is gimbal lock? not single solution with gimbal lock?
4. axis-angle $< - >$ rotation matrix - eigen basis, null space finding
5. Code/Visualize/Animate the math
6. how to tell which system has how many and what type of solutions? and what is a solution exactly?
7. sensing: hw1 p1 how is orientation not determinable?
8. $a^T * b = b^T$ * a iff a, b are vectors
9. what is determinant of transform matrix with only rotation and translation ? == 1
10. composition of linear transformations
11. rotation & quaternions

infinitesimal calculus

TODO

1. limits
2. the beauty of the phrase exact approximation
3. partial derivatives
4. derivative of vectors and matrices, jacobians, hessians, laplacians etc...

Statement 62 (Derivative).

Description: The derivative of a [function] $\Lambda : D \to C$ at $\alpha \in D$ is the limiting value of

$$\frac{\Lambda(\alpha + h) - \Lambda(\alpha)}{h}$$

as h tends to zero. Denoted by

$$\frac{d\Lambda}{d\alpha} \equiv \lim_{h \to 0} \frac{\Lambda(\alpha + h) - \Lambda(\alpha)}{h}$$

Significance: A notion of difference in function given difference in input. Generally visualized as a slope of the curve defined by the function. It is a general misconception that derivative is something very complicated and uncomprehensible thing. But it is actually not. It is a simple thing.

One subtle but important thing to notice here is that, there is no usage of the term "infinitesimals" here. The spirit of the derivative is in measuring change in function when input is changed. And the actual derivative itself is the value that the ratio approaches as the change gets smaller and smaller.

It can definitely be the case that the ratio does not have a limiting value. The derivative is not defined for such cases, that's all.

Proof. Proof by definition □

1 parents Function,

1 children Fundamental theorem of Calculus,

Statement 63 (Integral).

Description: The integral of a [function] $\Lambda : D \to C$ at $\alpha \in D$ is the limiting value of

$$\sum_{i=0}^{n-1} \Lambda(\alpha * \frac{i}{n}) * \frac{\alpha}{n}$$

as n tends to infinity. Denoted by

$$\int_0^\alpha \Lambda(\beta)d\beta \equiv \lim_{n\to\infty} \sum_{i=0}^{n-1} \Lambda(\alpha * \frac{i}{n}) * \frac{\alpha}{n}$$

Significance: A notion of sum of function until a given input. Generally visualized as area under the curve defined by the function. It is a general misconception that integral is some other worldly incomprehensible thing. It is not. It is just this. There is nothing more to the definition.

One thing to observe is that there is no method to natively evaluate the integral contrary to derivatives (which uses limits).

Proof. Proof by definition □

1 parents Function,

1 children Fundamental theorem of Calculus,

Statement 64 (Fundamental theorem of Calculus).

Description: This theorem links [derivative]s and [integral]s.
For [function]s $g, f : D \to C$ s.t. $\forall \alpha \in D$,

$$f(\alpha) \equiv \int_0^\alpha g(\beta)d\beta \implies \frac{df(\alpha)}{d\alpha} \equiv g(\alpha)$$

$$\frac{df(\alpha)}{d\alpha} \equiv g(\alpha) \implies f(\alpha) \equiv f(0) + \int_0^\alpha g(\beta)d\beta$$

.

Significance: Also provides a method to evaluate integrals. There is no native method to evaluate integral other than to basically guess viz. guess a function which upon derivation gives the original function and get its value at required point.

It is almost hilarious. It almost seems like the integrals are created like a hack. This should be developed and we should make a way to natively evaluate integrals that does not involve guessing.

Proof. The term 'approximation' used here is more profound than you might think.
First
If f is the integral function i.e. the area under the curve, consider a small finite change in the area df. Then

$$df \approx g(\alpha) * d\alpha$$
$$\frac{df}{d\alpha} \approx g(\alpha)$$

As the $d\alpha$ tends to 0

1. The limiting value of the LHS is derivative of f
2. The limiting value of the RHS is $g(\alpha)$
3. The approximation becomes equivalence

Therefore,

$$\frac{df}{d\alpha} \equiv g(\alpha)$$

Second
If g is the derivative of f

$$g(\beta) * d\beta \approx f(\beta + d\beta) - f(\beta)$$

$$\sum_{0}^{n-1} g(\alpha * \frac{i}{n}) * \frac{\alpha}{n} \approx \sum_{0}^{n-1} f(\alpha * \frac{i+1}{n}) - f(\alpha * \frac{i}{n})$$

Expanding and cancelling terms on RHS we have

$$\sum_{0}^{n-1} g(\alpha * \frac{i}{n}) * \frac{\alpha}{n} \approx f(\alpha) - f(0)$$

As n tends to infinity

1. The limiting value of the LHS is derivative of $\int_0^\alpha g(\beta)d\beta$
2. The limiting value of the RHS is $f(\alpha) - f(0)$
3. The approximation becomes equivalence

Therefore,

$$f(\alpha) - f(0) \equiv \int_0^\alpha g(\beta)d\beta$$

$$f(\alpha) \equiv f(0) + \int_0^\alpha g(\beta)d\beta$$

□

3 parents Derivative, Integral, Function,

probability

Statement 65 (Experiment).

Description: A list of instructions which when followed produces an outcome.

Significance: Anything that has some non-zero number of outcomes is an experiment. The meaning of this is not just limited to experiments which involve physics or chemistry. We are only concerned with the outcomes and not the inputs. Note that, at a time only one of the several possible outcomes of an experiment can occur.

Proof. Axiom. □

0 parents

7 children Deterministic Experiment, Non-deterministic Experiment, Outcome Space, Event, Mutually Exclusive Events, Event Space, Stochastic Experiment,

Statement 66 (Deterministic Experiment).

Description: Let O_k represent the final outcome when an [experiment] is repeated k times, an experiment such that $O_1, O_2, O_3, ...O_n$ are pairwise identical [identity] $\forall n \in \mathbb{N}$ [natural numbers] is called a deterministic experiment.

Significance: Can be used to represent experiments like adding two numbers.

Proof. Proof by definition □

3 parents Experiment, Identity, Natural Numbers,

0 children

Statement 67 (Non-deterministic Experiment).

Description: Let O_k represent the final outcome when an [experiment] is repeated k times, an experiment such that $\exists (i, j) \mid O_i \not\equiv O_j$ [identity] is called a non-deterministic experiment.

Significance: Note that outcome can remain same for some repetitions of the experiement. Can be used to represent experiments like rolling a dice.

Proof. Proof by definition □

2 parents Experiment, Identity,

0 children

Statement 68 (Outcome Space).

Description: The outcome space of an [experiment] is the [set] of its all possible outcomes [universal set].

Significance: A term to refer to all possible outcomes collectively.

Proof. Proof by definition □

3 parents Experiment, Set, Universal Set,

4 children Event, Properties of probability, Total Probability Theorem, Bayes Theorem,

Statement 69 (Event).

Description: An event E of an [experiment] is a [subset] of [outcome space]. An event is said to have occured $\longleftrightarrow$ $\exists$ outcome $\in E$ that occured.

Significance: The naming here is bad. An event is nothing but a set of outcomes. An event Q occured is equivalent to saying that the experiment's outcome was one of outcomes in Q. That's it. Nothing more. The number of elements in an event can be 0, 1 or even total number of possible outcomes.

Proof. Proof by definition □

3 parents Experiment, Subset, Outcome Space,

9 children Mutually Exclusive Events, Event Space, Stochastic Experiment, Probability, Properties of probability, Independent Events, Conditional Probability, Total Probability Theorem, Bayes Theorem,

Statement 70 (Mutually Exclusive Events).

Description: Two [event]s A and B of an [experiment] are mutually exclusive events $\longleftrightarrow A \cap B \equiv \phi$ [null set].

Significance: For example in the experiement of throwing a dice the event of getting an even number and the event of getting an odd number are mutually exclusive events. Note that they are not independent events, even both words are eerily similar. Mutually exclusive mean that at max one can occur at a given time. Independent means both can occur at the same time. Mutually exclusive events can occur in non-stochastic experiements too. Independent events can not occur in non-stochastic experiements.

Proof. Proof by definition □

3 parents Event, Experiment, Null Set,

0 children

Statement 71 (Event Space).

Description: The event space of an [experiment] is the [set] of all possible [event]s of the experiement.

Significance: A term to refer to all possible events collectively

Proof. Proof by definition □

3 parents Experiment, Set, Event,

2 children Stochastic Experiment, Random Variable,

Statement 72 (Stochastic Experiment).

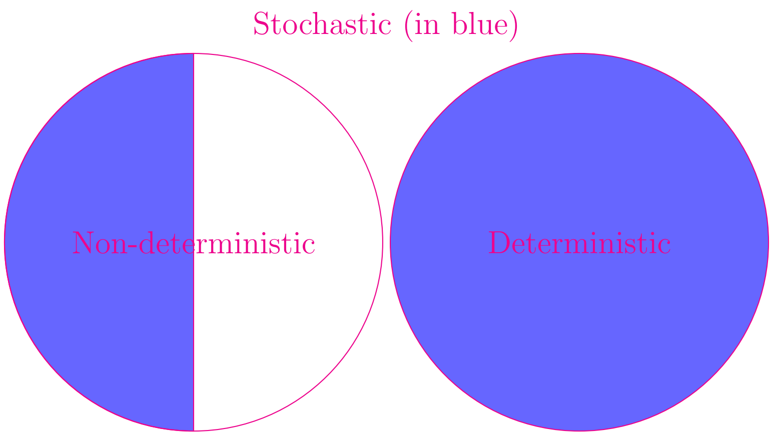
Description: An [experiment] such that $\forall$ [event] $\lambda \in$ [event space] the limiting value of the ratio

$$\frac{\text{number-of-times-}\lambda\text{-occured}}{\text{number-of-repetitions}}$$

as the number of repetitions tend to infinity, exists, is called a stochastic experiment.

Significance: Useful to build the concept of probability. Note that by my definition of stochastic experiment

1. All deterministic experiments are stochastic experiments with probability of the only outcome being 1.
2. Some non-deterministic experiments are stochastic experiments.



Proof. Proof by definition □

3 parents Experiment, Event, Event Space,

7 children Probability, Properties of probability, Independent Events, Conditional Probability, Random Variable, Total Probability Theorem, Bayes Theorem,

Statement 73 (Probability).

Description: The probability of an [event] λ of a [stochastic experiment] is the limiting value of the ratio

$$\frac{\text{number-of-times-}\lambda\text{-occured}}{\text{number-of-repetitions}}$$

as the experiment is the repeated infinitely many times. Denoted by $P(\lambda)$.

Significance: One might easily think that probability gives a sense of power of an event when an experiment is conducted. Although this is quite easy to misinterpret. A higher probability does not mean it will occur next. Even an event with unimaginabily miniscule probability may happen (many times in a row) and an event with very high probability may not occur (many times in a row). Say an experiement has two outcomes α, β with probabilities 0.999999999999 and 0.000000000001 respectively. But we can never say with certainity that when an experiment is conducted only α shall occur. Nor is the case when experiement is repeated say 100000000000 (any finite number) of times the corresponding ratios shall be 0.999999999999 and 0.000000000001, it can also happen that in all those repetitions α never occurs. This subtle nature of probability, that arises from the definition of limit itself, should be digested well.

A good way to remember this is to add meaning to notation. Consider notation,

$$P(\lambda)$$

Here

1. P should remind that probability is a limit definition.
2. ( ) contains an event(s) therefore an implicit experiment is defined. The experiment should be well understood. This is the most ignored step while working with probability.
3. λ represents an event, which is a fancy name for a set of outcomes of the experiement in context. The event occurs iff any of the outcomes in it occur.

Proof. Proof by definition □

2 parents Event, Stochastic Experiment,

3 children Independent Events, Conditional Probability, Random Variable,

Statement 74 (Properties of probability).

Description: Given S be the [outcome space] and A, B be some [event]s of a [stochastic experiment] E. The following properties hold.

1. $P(A) \geq 0$
2. $P(S) \equiv 1$
3. $A \cap B \equiv \phi \implies P(A \cup B) \equiv P(A) + P(B)$ [null set].

Significance: Some tools to played around with to create new tools.

Proof. From the ratio in the definition of probability, wheneven an experiment is conducted the numerator (num. times event E occured until now) ≥ 1 and denominator (num. of repetitions done) > 0

$\implies$ the ratio is always ≥ 0.

$\implies$ the limiting value of the ratio P(A) is also ≥ 0. When the event is identical to outcome space the ratio is identical to one.

$\implies$ the limiting value of the ratio P(S) is also $\equiv 1$. When the intersection b/w two events A and B is a null set, the number of occcurances of event $A \cup B \equiv$ number of occurances of A + number of occurances of B. Therefore the ratios at every repetition and thus limiting values of ratios are equal i.e. $P(A \cup B) = P(A) + P(B)$. □

4 parents Outcome Space, Event, Stochastic Experiment, Null Set,

0 children

Statement 75 (Independent Events).

Description: Two [event]s A and B of a [stochastic experiment] are independent $\longleftrightarrow P(A \cap B) \equiv P(A) * P(B)$ [probability].

Significance: Note that they are not mutually exclusive events, even though both words are eerily similar. Mutually exclusive mean that at max one can occur at a given time. Independent means both can occur at the same time. Mutually exclusive events can occur in non-stochastic experiements too. Independent events can not occur in non-stochastic experiements.

Proof. Proof by definition □

3 parents Event, Stochastic Experiment, Probability,

0 children

Statement 76 (Conditional Probability).

Description: Let P(B) denote [probability] of some non [null set] [event] B of a [stochastic experiment] E. Say event B occured

1. The probability that outcome a occured during the occurance of B is called conditional probability of a given B.

$$P(a \mid B) \equiv \frac{P(a \cap B)}{P(B)}$$

2. The probability that event A = $\{a_1, a_2, a_3..a_n\}$ occured during the occurance of B is called conditional probability of A given B.

$$P(A \mid B) \equiv \sum_{i=1}^{n} \frac{P(a_i \cap B)}{P(B)}$$

Significance: All it is saying is that one of some outcomes a.k.a B occured, what is the probability that during that occurance an outcome a occured.

Say the outcomes probabilities are $[a_1 = 0.1, a_2 = 0.2, a_3 = 0.3, a_4 = 0.4]$ respectively. Say event B $= \{a_1, a_2\}$ occured n times. To get $P(a_1|B)$. We take the ratio

$$\lim_{n\to\infty} \frac{\text{number-of-times-}a_1\text{-occured}}{\text{n}}$$

. The belief here is that at limit, a_1 occurs at same relative frequency (with other outcomes in B) when B occurs, as it did (with all outcomes of experiment) when the experiment occurs.

1. $a \cap B$ because if the outcome a is not in B, then it must definitely not have occured.
2. $\frac{1}{P(B)}$ because it normalizes each outcome's probability such that their sum $\equiv 1$.

Consider notation,

$$P(\alpha \mid \beta)$$

Here

1. P should remind that probability is a limit definition.
2. ( ) contains an event(s) therefore an implicit experiment is defined. The experiment should be well understood. This is the most ignored step while working with probability.
3. α, β each represents an event, which is a fancy name for a set of outcomes of the experiement in context. The event occurs iff any of the outcomes in it occur.
4. | should remind us that β occured.

Proof. Proof by definition ☐

4 parents Probability, Null Set, Event, Stochastic Experiment,

1 children Random Variable,

Statement 77 (Random Variable).

Description: A random variable of a [stochastic experiment] E is a [function] ψ mapping from E's [event space] S to $[0, 1]$ where

1. $\forall \lambda \in S, \psi(\lambda) \equiv$ the [probability] of the event λ.
2. $\forall \alpha, \beta \in S \mid \beta \not\equiv \phi, \psi(\alpha \mid \beta) \equiv$ the [conditional probability] of α given β,.

Significance: A wrapper around the definitions of experiment, event space, probability, conditional probability.

Proof. Proof by definition ☐

5 parents Stochastic Experiment, Function, Event Space, Probability, Conditional Probability,

2 children Total Probability Theorem, Bayes Theorem,

Statement 78 (Total Probability Theorem).

Description: For a [random variable] ψ of a [stochastic experiment] E, let B be some [event] of E and $\{A_1, A_2, ...A_n\}$ be a [partition] of [outcome space] of E, such that none of the A_i is a [null set] $\implies$

$$\psi(B) = \sum_{i=1}^{n} \psi(B \mid A_i)\psi(A_i)$$

Significance: Using this theorem we can split the probability of event B into conditional probabilities of B given A_is and proababilities of A_is.

Proof. **TODO** □

6 parents Random Variable, Stochastic Experiment, Event, Partition, Outcome Space, Null Set,

0 children

Statement 79 (**Bayes Theorem**).

Description: For a [random variable] ψ of a [stochastic experiment] E, let B be some [event] of E and $\{A_1, A_2, ...A_n\}$ be a [partition] of [outcome space] of E, such that none of the A_i is a [null set] $\implies$

$$\psi(A_i \mid B) = \frac{\psi(B \mid A_i)\psi(A_i)}{\sum_{i=1}^{n} \psi(B \mid A_i)\psi(A_i)}$$

Significance: Need for such a concept can be found at 3blue1brown and veritasium.

When you have a constant world but cannot be observed accurately and you can have evidence which determines something about the world, then you can update what you believe about the world by using this theorem.

Proof. **TODO** □

6 parents Random Variable, Stochastic Experiment, Event, Partition, Outcome Space, Null Set,

0 children

1. **TODO**
2. Cdf and pdf definitions
3. Expected value definitions
4. Joint probability definition
5. Joint vs Conditional probability
6. Functions of random variable
7. Random vectors
8. baysian networks, HMM, Markov, MDP, Q-learning ...
9. How to determine if an experiment is stochastic?
10. How to determine where the simple division conditional probability rule is not good?

fourier analysis

TODO

data analysis

TODO

1. data is a plural and a pronoun.
2. accuracy vs precision
3. estimation = estimate + uncertainity
4. modelling uncertainity using stochasticity
5. given a parametrized "stochastically noisy" model with stimulus and response (vector) variable and some samples, the process of estimating parameters (a.k.a fitting model). Using optimization for this estimation. Other non-optimization ways of doing the same. The essence of ML.

1. $(x_i - min)/(max - min)$ set of values vs normalizing set of values, show they are not equal in some instances and properties of each
2. emphasize diff b/w
   (a) min(sum(abs(residual))) ≡ min(Sum of Absolute Errors). The average of SAE is MAE
   (b) min(sum(residual * residual)) ≡ min(Sum of Sqaured Errors). The average of SSE is MSE
   (c) min(sum(perpendicular distance)) ≡ min(Sum of Perpendicular Errors). The average of SPE is MPE
3. Thinking mean and variance visually, mean of a list of numbers plotted on a coordinate system is "the rough line passing through the center", take differences from the rough center line to actual values and square them. Each of these towers or inverted towers can be thought to represent "certain variation" from the mean. Average (/n-1) of all this is defined as "variance". Etymology of "variance". observed the sqaured fn to ignore signs and preserve differentiability. Mind blown!!!. And all this from visualizing the definitions.
4. R-squared = Explained variation / Total variation. what it means (capturing the variation for just the samples in context and does not say anything about predictablility of model), and especially what it does not mean (good/bad model (patterns in residual plots)). Is it scale invariant?
5. Shallow vs deep classifiers ≡ seperate description and classification vs together description and classification.

1. does uncertainity change during fitting process?
2. patterns in residual plots as a measure of good or bad model.
3. Compare MLE vs MAP vs Linear Regression on.
   (a) fitting function
   (b) measure of accuracy, measure of uncertainity
   (c) covariance
4. Determining uncertainity after fitting in linear regression (or other models). Using info about training datum (a.k.a. sample) error term?

terminology

Statement 80 (Graphics).

Description: Graphics is an image on a 2D surface.

Significance: Often abused word so it is good to write down a proper definition of it. This is an uncountable noun, therefore the plural of graphics is also graphics. Graphics is by default singular. Examples of graphics are paintings on [paper, canvas, walls, solid object surfaces like balls, coffee mugs, stones], images generated on computer screen.

Proof. Axiom. □

0 parents

1 children Rendering,

Statement 81 (Rendering).

Description: The process of generating [graphics] from a 2D or 3D geometric scene is called rendering.

Significance: Often abused word, so good to jot down a proper definition. Graphics generated by computers are generally stored as a matrix of pixels.

It is also known as rasterization. But rasterization is often used to refer to a specific algorithm used for the process of rendering known as scan conversion, which is a misnomer. It is not a good practice. To refer to scan conversion, use the word scan conversion.

Proof. Proof by definition □

1 parents Graphics,

4 children Ray casting, Ray marching, Ray tracing, Scan conversion,

Statement 82 (Ray casting).

Description: Ray casting is an type of [rendering] in which rays are projected from camera into the scene through each pixel on image to detect intersections with objects analytically (a.k.a one form). These intersections are used to deterimine the color at each pixel.

Significance: A simple and useful form of rendering.

Proof. Proof by definition □

1 parents Rendering,

1 children Ray tracing,

Statement 83 (Ray marching).

Description: Ray marching is an type of [rendering] in which rays are projected from camera into the scene through each pixel on image in finite incremental steps. Intersections with objects are checked at each step. This is called a numerical approach (a.k.a many form). These intersections are used to deterimine the color at each pixel.

Significance: A simple and useful form of rendering.

Proof. Proof by definition □

1 parents Rendering,

1 children Ray tracing,

Statement 84 (Ray tracing).

Description: Ray tracing is a type of [rendering] in which either [ray casting] or [ray marching] is recursively applied to determine the color at each pixel.

Significance: A composition of ray casting / ray marching.

Proof. Proof by definition □

3 parents Rendering, Ray casting, Ray marching,

0 children

Statement 85 (Scan conversion).

Description: Scan conversion is a type of [rendering] in which objects in the scene are culled, clipped and transformed w.r.t a viewing fustum. Then each pixel corresponding to each object is colored.

Significance: Often ignored and/or badly replaced with the word rasterization. The operations being independent makes the process parallelizable. Actually this idea along with the motivation for making better graphics for games, movies, research etc... resulted in development of hardware that can run same instruction on multiple data simultaneously, which we now call as GPUs (graphics processing units). Although GPUs were invented for graphics other uses were discovered recently, a notable example being the use in parallelizing matrix computations for deep learning. GPUs that can be used for general purpose computations are named, surprise surprise, GPGPUs (general purpose GPUs).

OpenGL is made for scan conversion rendering only. It is not made for general purpose parallel computing, not even for other rendering techniques like ray tracing. OpenCL and CUDA are made for general purpose parallel computing. The purpose of OpenGL is not to get access to GPU, the purpose is rendering using scan conversion. The GPU just happens to be a good hardware for rendering using scan conversion.

Proof. Proof by definition □

1 parents Rendering,

languages

1. C/C++/Rust
2. Java/Kotlin/Go
3. Python
4. Shell

nursery

TODO: add some structure

Irony of thoughts

1. There exist things which we don't know but can approximate (ex. integrals)

Long project lessons

1. The best task management system until now is a white board and a marker
2. The best information note taking system is taking no notes at all
3. The best running notes are the ones that are forgotten shortly later
4. The best information gathering system until now is this journal

Fuzzily separated tree searches

1. Ones that have no state
2. Ones that have a fixed size state (finding nearest vertex to a position)
3. Ones that have a per-vertex state (path from start to current vertex)

Misc

1. Bitwise operations are much faster than arithmetic operations. So use them if possible. For ex in bit manipulation n, n - 1 bit manipulation trick
2. Linked list dummy initial node :P
3. Systemetic case expansion
4. Working with examples
5. Arrays that index main arrays to do things :P
6. Sometimes things to be done in second loop can be done directly in first

language

TODO: add some structure

1. Statically typed vs Dynamically typed == Compile time vs Run time type inference.
2. Compiled vs Interpreted == Transforming code to target machine code before running vs No transformation before running. Compiled might be faster because
   (a) It is not reading the code and transforming it for every line.
   (b) It can do global optimizations which might not be possible in interpreted as its field of view is local.
3. Statically vs Dynamically typedness is property of language. Compilation vs Interpration is the property of implementation not language; saying python is an interpreted language is wrong. Any language can be compiled. Any language can be interpreted.
4. Define, map and visualize the terms: Memory safety, Memory containment and Type safety.